(2)

REPORT DOCUMENTATION

AD-A255 281

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	
	May 1983	Unknown

4. TITLE AND SUBTITLE

Rational Belief

5. FUNDING NUMBERS

DAAB10-86-C-0567

6. AUTHOR(S)

Kyburg, Henry E., Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Rochester
Department of Philosophy
Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER

URCS - 3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army CECOM Signals Warfare Directorate
Vint Hill Farms Station
Warrenton, VA 22186-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

92-TRF-0045

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Statement A: Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)* There is a tension between normative and descriptive elements in the theory of rational belief. This tension has been reflected in work in psychology and decision theory, as well as in philosophy. Canons of rationality should be tailored to what is humanly feasible. But rationality has normative content as well as descriptive content. A number of issues related to both deductive and inductive logic can be raised. Are there full beliefs, statements that are just categorically accepted? Should statements be accepted when they become overwhelmingly probable? What is the structure imposed on these beliefs by rationality? Are they consistent? Are they deductively closed? What parameters, if any, does rational acceptance depend on? How can accepted statements come to be rejected on new evidence? A systematic set of answers to these questions is developed on the basis of a probabilistic rule of acceptance and a conception of interval-valued logical probability according to which probabilities are based on known frequencies. This leades to limited deductive closure, a demand for only limited consistency, and the rejection of Bayes' theorem as universally applicable to changes of belief. It also becomes possible, given new evidence, to reject accepted statements.

14. SUBJECT TERMS

Artificial Intelligence, Data Fusion, Rational Belief.

15. NUMBER OF PAGES

61

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to **stay within the lines to meet optical scanning requirements.**

Block 1. Agency Use Only (Leave blank).

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - DOD - Leave blank.
DOE - DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - NASA - Leave blank.
NTIS - NTIS - Leave blank.

Block 13. Abstract. Include a brief (Maximum 200 words) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (NTIS only).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

COGNITIVE SCIENCE TECHNICAL REPORT

University of Rochester
Rochester, NY 14627

URCS - 3
May 1983

RATIONAL BELIEF

by

Henry E. Kyburg, Jr.
Philosophy Department
University of Rochester

Rational Belief (Revised 9/6/82)

Abstract I

Henry E. Kyburg, Jr.

There is a tension between normative and descriptive elements in the theory of rational belief. This tension has been reflected in work in psychology and decision theory, as well as in philosophy. Canons of rationality should be tailored to what is humanly feasible. But rationality has normative content as well as descriptive content.

A number of issues related to both deductive and inductive logic can be raised. Are there full beliefs -- statements that are just categorically accepted? Should statements be accepted when they become overwhelmingly probable? What is the structure imposed on these beliefs by rationality? Are they consistent? Are they deductively closed? What parameters, if any, does rational acceptance depend on? How can accepted statements come to be rejected on new evidence?

Should degrees of belief satisfy the probability calculus? Does conformity to the probability calculus exhaust the rational constraints that can be imposed on partial beliefs? With the acquisition of new evidence, should beliefs change in accord with Bayes' theorem? Are decisions made in accord with the principle of maximizing expected utility? Should they be?

A systematic set of answers to these questions is developed on the basis of a probabilistic rule of acceptance and a conception of interval-valued logical probability according to which probabilities are based on known frequencies. This leads to limited deductive closure, a demand for only limited consistency, and the rejection of Bayes' theorem as universally applicable to changes of belief. It also becomes possible, given new evidence, to reject previously accepted statements.

1. Introduction.

The Greek philosophers conceived of man as a rational animal. This rationality was a potentiality that might or might not be actualized under certain circumstances. (The competence/performance distinction is hardly new.) Later philosophers have sought both to articulate the canons of rationality, and to apply them in an effort to understand both man and his world. In the course of Western philosophy, it has become ever clearer that the notion of rationality itself is problematic.

Hume divided the objects of knowledge into matters of fact and relations of ideas. Canons of rationality concern relations of ideas. In modern terms, these canons concern the logical relations among sentences or propositions. Whitehead and Russell, in their monumental treatise on mathematical logic, Principia Mathematica, (1959, 1910) attempted to provide a complete characterization of these canons. No sooner had they done so than objections arose. These objections fell into two groups: First, it was objected that many of the inferences allegedly licensed by the logical framework of the Principia were intuitively invalid; for example, from the denial of "If Road-Runner wins the fifth race, Speedy will not win the third," it might be alleged to follow both that Road-Runner wins the fifth and that Speedy wins the third. The canonical use of "$\supset$", "$\sim$", "$\wedge$", etc., does not precisely reflect the use of the corresponding English connectives. Second, it was objected that many intuitively valid inferences in ordinary language could not be captured in the formalism of Principia Mathematica, and this has led philosophical logicians to devise a plethora of modal, intensional, causal, and deontic logics.

Hume also emphasized the gulf between "ought" and "is." This comes to us as the injunction not to confuse the normative and descriptive.

The constraints on our beliefs -- on the relations of our ideas -- imposed by logic are intended to be a priori and normative, despite the fact that in designing those constraints we are guided by intuition and ordinary usage. But if we are going to be guided by ordinary usage, it behooves us to find out what that ordinary usage is -- clearly a task for empirical investigation rather than for armchair speculation. In recent years a number of psychologists (Johnson-Laird (1977), Henle (1962), Wason (1977)) have explored the inferential propensities of human subjects, and have discovered a considerable gap between the canons of rationality as codified in logic texts, and the ways in which ordinary people reason.

Does this show that people are not rational? Or that if they are potentially rational, they too rarely actualize that potentiality? Or that their performance falls short of their competence? Or does it show that logicians have formulated canons that do not, after all, capture the essence of human rationality?

Hume himself had few doubts about the nature of deductive relations. But he emphasized, more sharply than any of his predecessors, the difficulty of finding rational constraints for nondeductive argument. Scientific inference, learning from experience, probable argument, all escape the net of deductive rational constraints. John Maynard Keynes(1952, 1921) and Rudolf Carnap (1950) were among those to propose that there was a logical notion of probability that could be called on to provide rational constraints for nondeductive argument and inference. Such a notion would also provide a framework for decision theory. The program of finding rational constraints on nondeductive inference in the calculus of probability has not been a success. With regard to scientific inference, induction, and probable argument, many philosophers now argue that the quest for rational

constraints is misguided. Rather, we should look for historical, sociological, and psychological accounts of why people accept the arguments they do, make the inferences they do, believe the scientific theories and hypotheses they believe.

Probability plays a large role in philosophical discussions of decision making, choice, and belief, but the interpretation of probability most commonly employed is subjective: probability just is degree of belief. There is still a normative element: degrees of belief ought to satisfy the probability calculus, But in point of fact, this constraint is a purely deductive one. Combined with a behavioral interpretation of belief, it says roughly that you shouldn't be prepared to make a set of bets such that a wily opponent can be sure of taking from you what you value.

But again, if we suppose that probabilities reflect beliefs, and that behavior is a way of getting at beliefs, it becomes an empirical question whether people behave in the ways in which philosophers regard as rational. Psychologists have examined choice behavior, decision making, and the behavioral manifestations of degrees of belief (Edwards (1954), Tversky and Kahneman (1974), Nisbett and Ross (1980)), and discovered that people do not choose, decide, and believe as philosophical canons of rationality suggest they ought.

Again we are faced with a problem. Are the canons of rationality embodied in ordinary decision theory wrong? Or inappropriate for human beings? Or are people mostly irrational? Is there some way of adjusting the canons of rationality, or reinterpreting the actualities of behavior, so that the gap is not so great between what is and what ought to be? Or some way of modifying behavior to that same end?

These questions, and those raised previously, suggest certain prior questions. What is it that we want of a normative theory of rational belief? What sort of framework of terms and ideas should such a theory be placed in? What relation should we expect to find between a normative and descriptive theory of inference and choice? In the sections that follow, I shall try to provide an epistemological framework in which to seek answers to both normative and descriptive questions about belief, inference, and choice.

2. Methodology.

There are a number of possible sources of principles of rationality. In traditional philosophy, the source has often been taken to be rational intuition -- a faculty common to all men in virtue of their humanity, since men are rational animals. Even if there is such a faculty (which seems doubtful) the recent debates and disagreements concerning the nature and extent of constraints on rational belief show that it does not provide a univocal standard that can lead scholars to agreement.

L. J. Cohen (1981) suggests that the source of our standards of rationality is intuition -- the untutored intuition of ordinary educated people -- subject to the constraint of consistency. We should make the minimum modification in the deductive intuitions of the ordinary citizen to render those intuitions consistent. (Since consistency is itself a notion of deductive logic, the constraint seems either vacuous or question-begging.) This suggests that we should begin with an empirical inquiry into people's logical intuitions. But it is not clear that such an inquiry would be any more relevant to the development of normative standards of (inductive or deductive) logical cogency, than an inquiry into people's arithmetical intuitions would

be to the development of standards of arithmetical validity.

Stich and Nisbett (1980) also point to intuition as the grounds of justification in human reasoning, but suggests that it is the intuition of "experts" to which we should turn. They admit that "it is to be expected that there will be some disputes over justification that admit of no rational resolution," [p. 202] since my expert may be your crackpot. But this is precisely one of the things we would like a theory of rational belief to provide: a standard for sorting expert sheep from crackpot goats. Einhorn and Hogarth (1981) refer to "the inescapable role of intuitive judgment in decision making." (p. 61) Brian Ellis (1979) argues that the laws of belief <u>are</u> the laws of thought, though the laws of thought that interest us are the laws of ideal thought -- so to speak, the laws of frictionless thought, by analogy with the laws of frictionless billiard balls. These laws are discoverable by introspection, which I take to be roughly the same as intuition. But Ellis also argues that this is not true of the dynamic laws of belief -- the laws of changes of belief.

On the psychological side, there is a wide spectrum of empirical studies. Johnson-Laird (1977) contrasts the standard logical use of the truth-functional connectives and quantifiers with the use of the (allegedly) corresponding English constructions, and finds wide discrepancies. Kahneman and Tversky (1973) claim to show that people are often not "rational" in their assessments of probability. Slovic, Fishhoff, and Lichentenstein (1977) claim that "people systematically violate the principles of rational decision making." Mynatt, Doherty, and Tweeny (1977) have investigated "confirmation bias" in the assessment of scientific evidence. Lyon and Slovic (1976)

claim that people often fail to take account of base rates in making probability assessments.

There are several difficulties that stand in the way of taking these investigations to be immediately relevant to the investigation of canons of rationality. First, there is the problem of translation: "If P then Q" in English may have a truth-functional meaning, but it is more likely to mean one of (a) Q is derivable from P, (b) Q is derivable from P together with some other things I know, (c) P, together with other things I know, makes it probable that Q, (d) P, together with other things that I take experts to know, would render Q very probable. Perhaps there are other candidates as well. The point is that the standards of probabilistic and deductive cogency that are considered are abstract and formal; the material of an experimental investigation is necessarily concrete and framed in ordinary rather than formal discourse.

Another difficulty, emphasized by Einhorn and Hogarth (1981) in their brilliant review of psychological decision theory, is that the standards of rationality themselves are in dispute, so that it is unclear, when the intuitions of experimental subjects disagree with the intuitions of the experimenter, whether it is the experimenter or the subject who ought to reform his ideas of rationality.

Nevertheless, there are some clear cases -- among them some of those investigated by Kahneman and Tversky -- in which the subject himself seems likely to agree that he has made a "mistake." I have in mind particularly the example (Tversky and Kahneman (1981), p. 454) in which a subject prefers a sure gain of $240 to a 25% chance to gain $1000 and a 75% chance to gain nothing, but also prefers a 75% chance to lose $1000 and a 25% chance to

lose nothing, to the alternative of a sure loss of $750. When the two decisions are explicitly combined to yield a choice between a 25% chance to win $240 and a 75% chance to lose $760, against a 25% chance to win $250 and a 75% chance to lose $750, the second alternative is chosen by 100% of the subjects.

It seems, then, that some intuitions are pretty dependable and pretty universal. Nevertheless, in pursuing the consequences of even very simple intuitions, which constitute our starting point, we must be prepared to reexamine them at any juncture.

There is another respect in which intuition provides a starting point. In developing a theory, whether it is a normative one or a descriptive one, we must choose a representation for the domain with which we are concerned: the theory will concern certain objects, and certain relations among objects. We intend the theory, whether it is normative or descriptive, to account for or to influence a certain realm of experience. It may be that our choice of objects and relations is a poor one; that no theory framed in those terms can account for the realm of experience at issue, or that there is no way of applying a normative theory framed in those terms. Under these circumstances, the theory is not false, the standards not"incorrect" -- it is rather the case that the theory is simply ill formed.

The method that I shall follow, then, is primarily philosophical. I shall propose certain objects and certain relations as the ingredients of a theory of rational belief. I shall, on the basis of elementary intuitions, claim that certain of these relations actually hold of ideal beliefs, and that the resulting theory provides a rough approximation to actual human bodies of belief, and that, in addition, it can function

normatively, as what Ellis calls a "regulative ideal." For a regulative ideal, we require a theory that gives us a standard for the criticism and improvement of our actual beliefs (the suggestion that we believe only what is true does not give such a standard); but we also require a theory that goes beyond what we -- or even the experts -- actually do. It does no harm to have an ideal that can only be approached.

At the same time, empirical data are relevant. There is a wealth of data purporting to show that normative inference theory and normative decision theory are systematically and pervasively violated by actual human inference and decision making. Certain of these studies will be reviewed, and it will be suggested that in some cases and in some degree the normative theory sketched here is not so often or so flagrantly violated. This will be taken to show that a restructuring and modification of the classical normative theory of rational belief may lead to new questions and new research in psychological theory, as well perhaps, as to new approaches to the improvement of human performance.

There is another way in which empirical data can provide a "test" of a normative theory. We do not expect our subjects to conform completely to the normative theory. But when they fall short, we expect them to fall short in understandable ways. We expect normal adults, to whom our standards of rationality are applicable, to know the product of 4 and 7; we do not expect them to know the product of 56934 and 45927. For much the same reasons, we expect them to know the probability of a pair of ones resulting from the throw of a pair of dice, but we don't expect them to know the probability that two people in a group of 25 will have the same birthday. Nor do we expect them to be altogether accurate in intuitive

statistical inference.

Since our focus is /the question of belief and what makes it rational, we leave to one side all the very complex questions concerning utility that are involved in general decision theory. We shall be concerned with decisions only insofar as they throw light on actual beliefs.

Our procedure is thus intuitive: we shall seek principles of rationality that are intuitively valid in simple cases, and extend them to take account also of more complicated cases. At the same time we will look to empirical data to check our intuitions -- and also from another slant: Are failures to conform to the canons of rationality <u>understandable</u> in terms of natural human limitations?

3. The Terms and Scope of the Theory.

The objects of belief, those things to which an individual stands in a certain relation when he has beliefs, have been variously taken to be propositions, facts, states of affairs, sentences in mentalese, sentences in the ordinary language of the individual. I shall take them to be sentences in <u>our</u> language, but not our ordinary language. Rather, I shall suppose that they are sentences in an extensional first order logic, with the standard sentential connectives and quantifiers. This generates a challenge of interpretation. If someone / says, or acts as if he believes, that Rover is a dog, we can represent what he believes as the sentence "dog(Rover)," or "D(r)." But if someone says that John will get five dollars if he cuts the lawn, it is only when the speaker is a (slightly perverse) logician that we can sensibly represent his assertion as "$C(j) \supset F(j)$."

The burden of interpretation is nontrivial, particularly if, as I suspect, many conditionals are best interpreted metalinguistically, and not

by any sentence of our object language at all. (The claim that if P then Q is then interpreted along the lines: if sentence P is added to my body of knowledge (or to that of any sensible citizen), then the sentence Q will also be a part of that body of knowledge.) But there are also significant benefits conferred bythis move. We can characterize a demonstrably sound notion of logical validity. We can give a syntactical notion of proof from premises and of theoremhood. If you claim that A follows from B, and I am skeptical, then if we can agree on translations of A and B, T(A) and T(B), the issue can be resolved by a proof or a counter-example. And if we can't agree on translations of A and B, that in itself may help to show us where our differences lie.

Not everything we believe can be represented in a standard first order language. Certainly not everything that is believed can be represented in a first order language that we can understand: We can no more adequately represent a dog's beliefs in our regimented first order language than a treatise on painting can be translated into a language lacking color words. Our language is too poverty-stricken when it comes to odor words to do justice to the dog's beliefs. But there is still a wide range of things believed that can be represented in the way suggested, and a theory of rational belief, even if it were limited in scope to those things, would be interesting and useful.

In sum: we take objects of belief to be sentences in an extensional first order logic, with operations and identity, that includes the relation 'ε' (is a member of), and axioms for set theory.

In that language we have formal notions of deductive consequence (we write C is a deductive consequence of $P_1,\ldots,P_n$ as $\{P_1,\ldots,P_n\} \vdash C$),

consistency (we write Consis(S) for "the set of sentences S is consistent), and theoremhood (we write $\vdash$ C for "C is a theorem"). Note that these are purely syntactical notions; they rest on a syntactical notion of provability, rather than on a semantical notion of entailment.

There is a fundamental ambiguity in the notion of belief. We speak both of degrees of belief, and of belief simpliciter. When a coin is tossed, I believe it will come to rest (rather than going into orbit or disappearing), and I have a degree of belief of about a half that it will come to rest with the heads uppermost. One might try to collapse these notions: to construe full belief, or belief simpliciter, as belief of the highest degree or belief of degree 1 (Jeffrey 1965). Alternatively, we may construe belief simpliciter as acceptance into a set of statements that constitutes a body of knowledge or a rational corpus (Levi 1980). This is not the place to review the philosophical pros and cons of the two approaches to full belief. Perhaps the most penetrating discussion is in Levi (1980). I shall simply adopt the second approach.

Given that we adopt the second approach, there are a number of questions we should expect a theory of rational belief to answer. What is the structure of the set of statements constituting a rational corpus? How do statements get into a rational corpus? A less frequently addressed, but equally important question: How do statements get expunged from a rational corpus? Is what is taken to belong to a rational corpus dependent on context, and if so, in what way?

Degrees of belief are typically associated with probabilities. There are a large number of interpretations of probability available, including finite and limiting frequency interpretations (Russell (1948), von Mises (1957),

Reichenbach (1949)), various forms of propensity interpretations (Popper (1957), Mellor (1971)), logical range interpretations (Carnap (1950), Hintikka (1965)), subjectivist or personalist interpretations with varying degrees of normative force (Savage (1954), de Finetti (1980), Jeffrey (1965)) as well as a number of nonstandard views represented by Cohen's "Baconian" probability (Cohen (1977)), Popper's degree of corroboration (Popper (1959)), and various notions of "degree of factual support" (Hempel and Oppenheim (1945)), which do not satisfy the usual probability axioms.

I shall relate both degrees of belief and the grounds of acceptance into a rational corpus to my own interpretation of probability (Kyburg (1961), (1974)). This interpretation is syntactical, and probability is construed as a syntactical relation between a sentence and a set of sentences construed as a rational corpus. Given a sentence P, and a set of sentences S, if S meets certain minimal requirements, the probability of P relative to S will be a closed subinterval [p,q] of the interval [0,1].

An example may help to clarify the definition that follows. Suppose that P is the sentence "John will go to the movies tonight," and that S is the set of sentences that represent my reasonable or justified beliefs this afternoon. The probability, relative to my body of knowledge S, that John will go to the movies tonight is the interval [.6,.7] under these circumstances: first, S represents a set of reasonable beliefs. Second, I know that John is going to decide whether or not to go to the movies by drawing a chip from an urn, and that he will go if and only if the chip he draws is black. If we represent the proper description of that chip by c and the set of black objects by b, the sentence "$P \equiv \overset{c \in}{\longleftarrow} b$" is among the sentences representing my reasonable beliefs. Third, I know that the chip c is a member of the set of chips in the urn; representing this set by u,

the sentence "$\underline{c} \in \underline{u}$" also belongs to $\underline{S}$. Fourth, what I know about the set of chips $\underline{u}$ is that between 60% and 70% are black, and I don't know anything more exact than that. The sentence expressing this fact will be written "$\%(u,\underline{b}) = [.6,.7]$"; we suppose that "$\%(u,b) \in [.6,.7]$" is a member of $\underline{S}$. Fifth, relative to what I know -- $\underline{S}$ -- the chip $\underline{c}$ must be a random member of the set of chips $\underline{u}$ with respect to being black. This is taken to require, not that I have some special knowledge about how $\underline{c}$ is selected, but that there are no ingredients in $\underline{S}$ which would lead me to a conflicting probability. For example, if I knew that John wanted to go to the movies very badly and that he was likely to peek at a chip before he "selected" it, I would have grounds for denying that $\underline{c}$, the chip selected by John, was a random member of $\underline{u}$ with respect to being black.

The definition of probability is roughly the following: The probability of the sentence $\underline{P}$, relative to the set of sentences $\underline{S}$, is the closed interval $[\underline{p},\underline{q}]$, $\underline{Prob}(\underline{P},\underline{S}) = [\underline{p},\underline{q}]$, just in case there are terms $\underline{x},\underline{y},\underline{z}$ such that:[1]

(1) $\underline{S}$ is a rational corpus. This requires that $\underline{S}$ satisfy certain syntactical constraints that will be discussed shortly.

(2) "$\underline{P} \equiv \underline{x} \in \underline{z}$" is a sentence in $\underline{S}$.

(3) "$\underline{x} \in \underline{y}$" is a sentence of $\underline{S}$.

(4) "$\%(\underline{y},\underline{z}) \in [\underline{p},\underline{q}]$" is a sentence of $\underline{S}$.

This last is a sentence saying that the proportion of $\underline{y}$'s that are $\underline{z}$'s lies in the interval $[\underline{p},\underline{q}]$. It may also be interpreted in terms of relative frequency, or limiting frequency, or, most generally, measure. Levi, whose interpretation of probability is quite close to mine, requires that it be a statement of chance (Levi (1980), p. 251); if we so interpret it, we must construe it nonextensionally.

(5) $\underline{x}$ is a random member of $\underline{y}$ with respect to $\underline{z}$, relative to the set of sentences $\underline{S}$.

This last condition is also to be spelled out syntactically. It is equivalent to the assertion that, relative to $\underline{S}$, $\underline{y}$ is an appropriate reference class for the question of whether $\underline{x}$ belongs to $\underline{z}$. To give a flavor of the syntactical constraints embodied in condition (5), we may note that if "$\underline{x} \in \underline{y}'$", "$\underline{y}' \subset \underline{y}$", and "$\%(\underline{y}',\underline{z}) \in [\underline{r},\underline{s}]$" belong to $\underline{S}$, where $[\underline{r},\underline{s}]$ is different from $[\underline{p},\underline{q}]$, then we will want to say that condition (5) is not satisfied. That is, we will deny (5) if $\underline{S}$ includes the knowledge that $\underline{x}$ is a member of some subset of $\underline{y}$ in which the relative frequency of $\underline{z}$'s differs from that represented by (4).

We may illustrate this special case by reference to the previous example. Suppose my body of knowledge or rational beliefs, $\underline{S}$, is expanded by the addition of sentences representing the knowledge that there are two kinds of chips in the urn, round ones and square ones; that in deciding whether or not to go to the movies John always chooses (deliberately) a round chip; and that between 15% and 20% of the round chips are black. As before, $\underline{c}$ is the chip that John will choose, $\underline{b}$ is the set of black objects, $\underline{u}$ the set of chips in the urn. But now we must consider the partition of $\underline{u}$ into two subsets: $\underline{ru}$ consisting of round chips and $\underline{su}$ consisting of square chips. We have

$$\underline{c} \in \underline{ru}$$

$$\underline{ru} \subset \underline{u}$$

$$\%(\underline{ru},\underline{b}) \in [.15,.20]$$

all in $\underline{S}$, and therefore we will deny that $\underline{c}$ is a random member of $\underline{u}$ with respect to $\underline{b}$, relative to the expanded $\underline{S}$. But we will now be able to claim that $\underline{c}$ is a random member of $\underline{ru}$ with respect to $\underline{b}$, and therefore

that the probability that John will go to the movies is the interval [.15,.20].

It is also necessary to impose constraints on the terms that can occupy the places of y and z in the definition. All of these matters are somewhat difficult, but they do not bear directly on the issues at hand. What I am supposing, and claim to have shown elsewhere, is that it is possible to give rules that pick out a reference class for a sentence P, given a set of sentences S representing a rational corpus.

The notion of probability that I have sketched has the following properties:

(P-1) If S is a rational corpus, and P and Q are known in S to have the same truth value (i.e., if "P ≡ Q ∈ S), then Prob(P,S) = Prob(Q,S).

In the modified example, since we still know in S that John will go to the movies if and only if he draws a black chip, we have:

Prob("John will go to the movies",S) =

Prob("John will draw a black chip",S) = [.15,.20].

(P-2) If S is a rational corpus and P ∈ S, then Prob(P,S) = [1,1] and Prob(~P,S) = [0,0].

In the modified example, since "ru ⊂ u" is in S, Prob("ru ⊂ u",S) = [1,1]. Since "c is round" is in S, Prob("c is not round",S) = [0,0].

(P-3) Every probability is based on a relative frequency known in S (or measure, or chance, or limiting frequency).

In the initial version of the example, the probability that John will go to the movies is based on the knowledge that %(u,b) ∈ [.6,.7]; in the modified example it is based on the knowledge that %(ru,b) ∈ [.15,.20].

(P-4) If $\underline{S}$ consists of a set of true sentences of the form "$\%(\underline{y},\underline{z}_i) \in [\underline{p}_i,\underline{p}_i]$," and a sentence "$\underline{x} \in \underline{y}$," the definition of probability is equivalent to a frequency definition correspondingly restricted in scope.

Suppose that "all we know" about $\underline{c}$ is that it is a chip in the urn, but we know of the set $\underline{u}$ of chips in the urn that there are four kinds, $\underline{k}_1$, $\underline{k}_2$, $\underline{k}_3$, $\underline{k}_4$, and that the relative frequency of each kind is 1/8, 1/4, 1/16, 9/16 = 1 - 1/8 - 1/4 - 1/16. The probability that $\underline{c}$ belongs to any kind (or any Boolean combination of kinds) is the corresponding frequency.

(P-5) If $\underline{S}$ is a rational corpus and $\underline{T}$ a finite set of sentences, there exists a function $\underline{B}$ (a belief function) such that $\underline{B}$ satisfies the axioms of the conventional probability calculus, and for every sentence $\underline{P}$ in $\underline{T}$, $\underline{B}(\underline{P}) \in \underline{Prob}(\underline{P},\underline{S})$

Suppose that $\underline{t} = \{\underline{t}_1,\underline{t}_2,\underline{t}_3\}$ and $\underline{Prob}(\underline{t}_1,\underline{S}) = [\underline{p}_1,\underline{p}_2]$, $\underline{Prob}(\underline{t}_2,\underline{S}) = [\underline{p}_2,\underline{q}_2]$, and $\underline{Prob}(\underline{t}_3,\underline{S}) = [\underline{p}_3,\underline{q}_3]$. Then there exists an additive real-valued function $\underline{B}$ such that $\underline{B}(\underline{t}_1) \in [\underline{p}_1,\underline{q}_1]$, and $\underline{B}(\underline{t}_3) \in [\underline{p}_3,\underline{q}_3]$.

(P-6) If $\underline{S}$ is a rational corpus, and $\underline{P}$ is derivable from $\underline{Q}$, then the lower probability of $\underline{P}$ is at least as great as the lower probability of $\underline{Q}$.

In the initial example, the probability that John will go out somewhere tonight is an interval whose lower bound is at least .6.

(P-7) The principle of epistemic conditionalization -- that if $\underline{P}$ is added to $\underline{S}$ to yield a new rational corpus $\underline{S}'$, then the probability of $\underline{Q}$ relative to $\underline{S}'$ should be the ratio of the probability of "$\underline{P}$ and $\underline{Q}$" to the probability of $\underline{P}$ (where "ratio" is suitably understood) does not generally hold.[2]

(P-8) If the probability of P relative to S is [p,q], the probability of ~P relative to S is [1-q,1-p].

In our original example, the probability that John will go to the movies is [.6,.7]. The probability that he will not go to the movies is [.3,.4] = [1-.7,1-.6].

The particular framework that I discuss here is not the only one of its kind. There are other, similar, ways of approaching these problems -- for example, Levi (1980) develops a view of rationality that is epistemic in character, and depends even more on the notion of commitment, but which is distinguished from the present account both by the importance of chance in his treatment of epistemic probability, and by a more thorough pragmatic orientation. On Levi's view the principle of epistemic or confirmational conditionalization does hold.

4. The Principles of Full Belief or Acceptance.

What is it to accept a sentence, or to award full belief to it? We might construe belief in this sense as occurrent belief: one is fully believing or accepting S only if one is thinking of it, and thinking of it in a certain way. But this would not lead us to a very interesting normative theory. We might construe belief dispositionally: X believes the sentence S if X is inclined to assent to it, if asked, or if X has a (non-probabilistic) disposition to act as if it were true. But again this seems not strong enough to give rise to an interesting normative account of rational belief.[3] I think of myself as accepting the axioms of set theory, for example, but there are many theorems of arithmetic that I am too ignorant to assent to. On the other hand, if I were offered a proof of such a theorem, I would not only be inclined to assent to it following

the proof, but feel that I had been committed to the acceptance of that theorem all along. When I accept the axioms of set theory, I am committing myself to accepting their consequences. Of course, if those axioms have untoward consequences -- for example, if they should prove to be inconsistent -- I will not then regard myself as committed to believing all the sentences of the language. Rather, were I to become aware that the axioms of set theory commit me to everything, I would no longer accept those axioms, but would rather seek out some fixed up set of axioms.

I shall construe full belief or acceptance as commitment. To accept P is to be committed to P, and also to its deductive consequences. But we may still ask how far this committment takes us: does it commit us to the consequences of the whole set of sentences we accept? Levi (1980) argues that it does. I would argue against this, not on grounds of human finitude or logical limitation -- that would speak against being committed to all the theorems of set theory -- but on the grounds that it does not seem plausible to demand that the set of empirical sentences we accept be deductively closed. Here are four examples:

(1) I can believe of each statement that I accept in a certain context that it is true (or else I wouldn't accept it) and also reasonably believe that/(some)one of them is false -- I can believe the negation of the conjunction of those statements.[4] The situation can be represented as

$$\underline{B}(p_1)$$
$$\underline{B}(p)_2$$
$$-$$
$$-$$
$$-$$
$$\underline{B}(p_n)$$
$$\underline{B}(\sim(p_1 \wedge p_2 \wedge \ldots \wedge p_n))$$

(2) The lottery problem is well known. If reasonable acceptance is to be grounded in probability alone, then one should believe of any specified ticket in a fair million ticket lottery that it will not win the grand prize. But the conjunction of a million such statements contradicts the assertion that the lottery is fair. Our epistemic state is highly symmetrical with respect to the tickets; the evidence that any given ticket will lose is overwhelming; the evidence that not all the tickets will lose is even more overwhelming. To accept some but not all the statements of the form "ticket i will not win" is grotesquely arbitrary. To reject them all, on the grounds of the very symmetry that leads to the problem is to give up too much; the same grounds would undermine the arguments used to justify "the rejection of the null hypothesis" in statistics. Each possible sample, in such an argument, is assumed to be "equally probable," and the null hypothesis is rejected exactly on the grounds that if the hypothesis were true, we would have to suppose that we had drawn the winning ticket, which is too improbable to be believed.

(3) In statistical inference, if we allow ourselves to speak of the probability of statistical hypotheses at all, there are cases in which many hypotheses have the same high probability, but in which their conjunction is inconsistent. For example, consider the inference from a sample of a normal population of unknown mean to a value for that mean. Given the sample mean, the probability that the unknown mean lies in any interval is given by a "fiducial" distribution, integrated over that interval. For any number $1 - \varepsilon$ less than 1, there are an infinite number of intervals such that the fiducial probability is $1 - \varepsilon$ that the unknown mean belongs to that

interval. But the intersection of these intervals is very small. We can also have a fiducial probability of $1 - \varepsilon$ that the mean does not belong to this intersection.

(4) In measurement theory we often suppose ourselves to be making a large number of direct comparisons of objects. Consider a set of n statements of the form "A_i is the same length as A_{i+1}." (The first board for our bookshelves is the same length as the second; the second is the same length as the third; ...) Clearly each of these statements may be acceptable when the statement "A_0 is longer than A_{n+1}" is also acceptable. (The first board is definitely longer than the last.) There are no grounds for choosing among them. But to reject them all is to give up on measurement. One might say that we should reject them all; that the lesson to be learned is that we cannot judge "equality of length" directly. But approximate equality of length won't give us measurement theory (for example, it isn't transitive).

How do sentences become accepted? One answer is: when they are probable enough. Suppose that a sentence P is so highly probable as to be practically certain, relative to a set of sentences S construed as a rational corpus. Do we want to regard P therefore as a member of S? It is awkward to do so, for then P is not probable or practically certain, relative to S, but (by P-2) certain, relative to S. Furthermore, we must face the problem of saying what degree of probability is required for acceptance in S. One way of approaching these problems is to distinguish two levels of rational corpora. Let S_e be the evidential corpus, and S_p be the corpus of practical certainties. We then adopt the following principle of rational acceptance:

Principle I: A sentence P is acceptable in the corpus of practical certainties indexed by the real number p, if and only if there is an evidential corpus S_e, indexed by a number e larger than p, such that the minimum probability of P relative to S_e is greater than p.

In virtue of P-6, this principle yields just the structure we have been talking about.

For example, in the affairs of ordinary life, we might take $\underline{e} = 0.999$ and $P = 0.99$. We can take it as evidence that a given coin is fair: that is, we can include a statement in our corpus of evidential certainties $\underline{S}_{.999}$ that the distribution of heads in tosses of <u>this</u> coin is approximately binomially distributed with a parameter close to a half -- say $.500 \pm .010$. Relative to $\underline{S}_{.999}$, then, the probability of heads on a single toss is $[.490,.510]$. But if we consider a sequence of seven tosses, the maximum probability of getting all heads is less than .01; the minimum probability of not getting seven heads in seven tosses is at least 0.99; and therefore we may regard it as <u>practically certain</u> that we will not get seven heads in the next seven tosses. "The next seven tosses will not all result in heads," will appear in $\underline{S}_{.99}$, our corpus of practical certainties.

How about $\underline{S}_{\underline{e}}$, the evidential corpus? If what is in $\underline{S}_{\underline{e}}$ is at issue, we construe the level of $\underline{S}_{\underline{e}}$ as practical certainty, and require that there be a proto-evidential corpus, $\underline{S}_{\underline{e}}'$ relative to which each ingredient of $\underline{S}_{\underline{e}}$ have a high enough probability. In other words, we would simply apply the same principle at a higher level.

In the previous example we may ask for the grounds on which we can accept in $\underline{S}_{.999}$ the assertion that the distribution of heads is binomial with a parameter of $.500 \pm .010$. To answer we would take $\underline{p} = .999$ and $\underline{e} = .999$ (say). We include in $\underline{S}_{.999}$ our knowledge of the dynamics of coin tossing and our knowledge of the design of coins and the procedures for minting them. The probability that for <u>this</u> coin the distribution of heads is approximately binomial with parameter $.500 \pm .01$ is at least .999 relative to $\underline{S}_{.999}$.

The value of p, the level of probability required for acceptance into S_p, concerns us both as a parameter in the normative theory of rational belief and as a parameter in the corresponding descriptive theory. Dretske (1982), among others, has argued that "there seems to be no non-arbitrary place to put a threshhold" (p. 8). There are two answers to this claim.

It seems inevitable that what will strike us as an appropriate level of p for one context will be inappropriate in another. For the normative theory, therefore, we need a way of classifying contexts. My suggestion is this: that we consider relatively global contexts, and characterize them in terms of the maximum or minimum odds at which our chancey decisions in those contexts might pay off. Thus in "ordinary life" we do not ordinarily consider gambles involving odds of greater than 20:1. This would suggest $p = 0.95$ as a suitable acceptance level. There are special contexts that we all face on occasion when this does not seem appropriate: when buying insurance for example. In that case we are contemplating a gamble in which the odds may be 100:1 or 1000:1 or greater. An appropriate level of practical certainty might then be .99 or .999. In scientific inquiry or in public policy, the stakes may be similarly extreme, and the level of acceptance may be similarly stringent.

At the other extreme, some conservative academic epistemologists seem to feel that we can avoid (or should) avoid chancey decisions. Since the maximum payoff you can offer is the reciprocal of the maximum payoff you can receive, this suggests that the range of odds contemplated in the epistemic context is close to 1:1; and this would lead to a value of p very close to (just over) 1/2. ("You can reasonably believe P if it is more probable than not.")

This is a relatively new question, and so far as I know relatively little thought has been devoted to it. But, as far as the normative theory is concerned, it certainly seems premature to claim that there is no non-arbitrary threshold.

With respect to the descriptive theory, we may be better off. The parameter p is an adjustable one whose value can be chosen to make the most sense of the data we have. Some indication that this might be a useful way to approach the data of behavioral decision theory will be given in a subsequent section.

There are a number of consequences of Principle I that are worth remarking.

(1) If a sentence P has a minimum probability greater than p relative to S_e (or less than $1-p$), it is not rational to bet against it (on it) at any odds in the range characterizing practical certainty. It is a practical certainty, a datum. But as the above discussion of the value of p suggests, if one is actually offered enormous odds, that in itself may suffice to change the context to one in which a different (larger) value of p is appropriate. In an ordinary context, I simply accept the statement that about half the tosses of this coin will land heads. But if the relative pay-off were great enough, I would consider a bet on the question of whether or not this coin is biassed -- i.e., on the truth of a statement that in another context I accepted as a datum.

(2) Approximate statistical statement[5] of the form "$\%(A,B) = [p,q]$" can be rendered probable enough for acceptance by observational evidence in S_e.[5] In its crudest form this inference makes use of (1) the set-theoretical statistical truth that whatever be the proportion of A's

that are B's the proportion of large subsets of A that have a frequency of B's close to that among A's in general is very large; (2) the sample of A's that we have observed is a random member of the set of those large subsets relative to what we know; and (3) the proportion in the sample, known to have 100% B's, differs by δ from the proportion among A's in general if and only if the general proportion lies in the interval $r \pm \delta$. Thus, (1) whatever be the proportion p of black balls among draws with replacement from urn u, the proportion of sets of 10,000 draws that have a relative frequency of black balls differing by less than .02 from p is at least .98; this sample of 10,000 has 40% black balls; this sample is a random member of the set of samples of 10,000; therefore the probability is [.98,1.0] that $p \in [.40-.02,40+.02] = [.38,.42]$. Note, however, that precise statistical statements of the form "$\%(A,B) \in [p,p]$" and in particular statements of the form "$\%(A,B) \in [1,1]$" corresponding to "All A's are B's") cannot in general be rendered highly probable by observational evidence.

(3) Statements that are characteristic of the language L -- i.e., statements that are tautologies in that language -- automatically receive probability 1 and are automatically accepted in S_p. Furthermore, given any such statement T, and any statement P probable enough to be accepted in S_p, their conjunction will be probable enough to be accepted in S_p, and therefore (by P-6) so will any deductive consequence of their conjunction.

(4) From (2) and (3) it follows that any universal generalizations -- "All A's are B's", for example -- that are in S_p are to be construed as tautologies of the language L, or else are to be construed as approximate statistical generalizations: "Almost all A's are B's".

(It is an *a priori* constraint on our language that all bears are mammals; it is an arbitrary rule of post office procedure that all sealed letters must have 6d stamps; but it is only "approximately true" that all people who go to three drugstores in a row have failed to find what they wanted in the first two. It is more accurate to say that it *is* true that *almost* all people...) As will become clear later, this is an important distinction, since "All A's are B's entails its contrapositive, "All non-B's are non-A's", while the approximate statement, "Almost all A's are B's" does not entail "Almost all non-B's are non-A's."

(5) It is possible to argue that since much of scientific theorizing is concerned with the establishment of truly universal generalizations, it is better to look on it in terms of the choice between languages characterized by different meaning postulates or tautologies, than as a matter of testing or attempting to falsify universal generalizations. If this is so, it may explain some of the difficulties surrounding attempts to explore or inculcate a Popperian approach to scientific inference.

(6) A deductive argument will show that the corpus S_p is committed to the statement P only when the *conjunction* of the premises of the deduction is in S_p. We require '$P_1 \wedge \cdots \wedge P_n$' $\in S_p$ and $\{P_1,\ldots,P_n\} \vdash P$, and not merely $\{P_1,\ldots,P_n\} \vdash P$.

A classic difficulty in the theory of knowledge has stemmed from the urge to regard observation (or perception) as an incorrigible foundation of knowledge, and at the same time to recognize that observation --

particularly in science, and particularly in the form of measurement -- must be regarded as fallible. It is clear that we must admit the results of observation into our evidential corpus if we are to learn from experience at all. Yet it is also clear that however confident we may be of an observation, there is a possibility (which often cannot simply be "ignored") that it is in error.

The framework already proposed suggests a resolution of this difficulty. In learning a language, whether ordinary language at the age of two, or the specialized language of livestock judging at the age of thirty-two, one is learning to make observational judgments (among other things). One also learns that one's judgments are fallible: one never achieves perfection. Errors may be pointed out by one's teachers, or they may become obvious through conflicts with other things -- including generalizations characteristic of the language -- one knows. In either event, one can learn from one's mistakes: one can learn (in a metalinguistic rational corpus) that observational judgments of kind $\underline{K}$ are wrong some small portion of the time: say about ε.

Consider a sentence $\underline{P}$ reflecting an observational judgment of type $\underline{K}$. If $\underline{P}$ is, relative to what one knows about it, a random member of $\underline{K}$, the probability that it is mistaken is about ε. If $1 - \varepsilon$ is less than $\underline{p}$, $\underline{P}$ may be accepted as practically certain -- as a member of $\underline{S}_{\underline{p}}$. Note that $\underline{P}$ may be a random member of $\underline{K}$ at one time, and cease to be a random member of $\underline{K}$ when new information relevant to the chance that $\underline{P}$ is mistaken becomes available.

In fact, since, by P-8, $\underline{P}(\underline{S},\underline{K}) = [\underline{p},\underline{q}]$ if and only if $\underline{P}(\sim\underline{S},\underline{K}) = [1-\underline{q},1-\underline{p}]$, it is perfectly possible to have

$\underline{P}(\underline{S},\underline{K}) = [\underline{p},\underline{q}]$: $\underline{S}$ is practically certain relative to $\underline{K}$

$\underline{P}(\sim\underline{S}, \underline{K} \cup \underline{K}') = [1-\underline{q}',1-\underline{p}']$: $\sim\underline{S}$ is practically certain relative to an expansion of $\underline{K}$ by $\underline{K}'$.

Even observational statements may come to be accepted and then come to be rejected in the light of new information on the model suggested. For example, one may accept on the basis of observation that there is a cat on the porch, and then, on closer investigation, which results in the addition of new evidential statements to one's rational corpus, be led to reject that statement and to accept the statement that it is not a cat (but, say, a mongoose) on the porch. That does not mean that one was not justified in first accepting that it was a cat. A more common example is to accept on the basis of measurement that the melting point of compound $\underline{X}$ is $\underline{k} \pm \underline{d}$ degrees, and then, on the basis of more, and more careful, measurements, to reject that first measurement as yielding an "outlier".

The principle that justifies this is essentially just a metalinguistic version of principle I. It is phrased in terms of error, and it refers to the evidential corpus $\underline{S}_{\underline{e}}$ rather than $\underline{S}_{\underline{p}}$ for convenience. Recall that both $\underline{e}$ and $\underline{p}$ are numbers adjustable to fit the context.

A sentence $\underline{P}$ is acceptable in the corpus of evidential certainties indexed by the real number $\underline{e}$ on the basis of observation if and only if the probability that $\underline{P}$ is mistaken, relative to the meta-evidential corpus $\underline{MS}_{\underline{e}}$, is less than $1 - \underline{e}$.

The meta-evidential corpus $\underline{MS}_{\underline{e}}$ embodies our knowledge about the frequency of mistakes among statements of various classes based on observation.

Thus Principle I provides for the acceptance of fallible observation statements. Technical object-language metalanguage complications arise in the general application of Principle I to both observational and directly statistical uncertainties. These complications have been discussed in some

detail in (Kyburg, 1983b); they have only a marginal bearing on the issues with which we are concerned here, and will not be further discussed.

Given the syntactical notions of provability and probability, then, and given a formal language L in which the beliefs of an individual can be expressed, we take Principle I to say all there is to say about full belief.

5. The Principles of Partial Belief.

The subjective interpretation of probability generally supposes that a degree of belief can be represented by a real number in the interval [0,1]. Various coercive procedures -- e.g., forced bets -- have been suggested as a way of determining the degrees of belief of experimental subjects. At the same time, some writers (Savage (1966), Jeffrey (1974), Good (1962)) have suggested that beliefs be characterized on more than one dimension: one might feel "less secure" about one's degree of belief of .3149725 that the next president will be a Republican, than about one's degree of belief of .3149725 that of the next 75 flips of this coin, between 39 and 35 will result in heads.

It has also been suggested (Smith (1961), Dempster (1968), Kyburg (1961), Good (1962), Shafer (1976), Levi (1974)) that one way to capture this dimension is through the use of intervals to represent degrees of belief. Smith (1961) suggests that a first approach to a behavioral counterpart of an interval of belief would be given by the least odds at which a person would be willing to bet on P, and the least odds at which he would be willing to bet against P. In general one might suppose that these odds are not reciprocals, and that they would define an interval. This approach is rather crude, and is excessively sensitive to the person's

interest in or distaste for gambling. A more accurate way of measuring intervals of belief would be of considerable interest.

Even without a way of getting at interval beliefs directly, however, we can apply the present theory. If, relative to an agent's body of practical certainties, the probability of the statement P is the interval [p,q], then it is clearly irrational of him to be willing to pay more than q dollars for a ticket that yields a dollar if P is true and nothing otherwise. Similarly, it would be irrational of him to sell a ticket for less than p dollars that obligated him to pay a dollar if P is true. Thus we can use many of the standard ways of getting at alleged real number degrees of belief to apply the present theory. We can put the matter more generally by saying that a person's behavior with respect to a sentence P should be compatible with his having a (hypothetical) degree of belief in P that falls within the interval representing the probability of P.

We state this as Principle II:

Principle II: If S_p is X's corpus of practical certainties in a certain context, then X's degree of belief in a sentence P, whether it is construed as an interval of the form [q,q], or as a non-degenerate interval, should be included in $\underline{Prob}(P,S_p) = [q,r]$.

To borrow from an earlier example, if relative to my body of evidential certainties, the probability that John will go to the movies tonight is [.6,.7], then my "degree of belief" that John will go to the movies tonight should be some number -- e.g., 0.62 -- in that interval, if we construe degrees of belief as measured by real numbers, or should be some subinterval of [.6,.7] -- e.g., [.63,.65] -- if we take "degrees of belief" to be better represented by intervals.

Finally, we need a principle connecting degrees of belief and utilities with actions. We adopt the principle of maximizing expected utility as our third principle of rationality. We assume that X has a real-valued utility function defined over statements. (This is a questionable supposition but one we shall not question here.) The outcomes of decisions or choices can be described as the coming true of certain propositions. Of course it is the utility of the total world outcome that concerns the actor, but in common with most writers, I shall assume that in artificially simplified situations we can often work with the marginal utilities associated with sentences describing partial outcomes: winning a dollar if the coin lands heads, and losing a dollar if the coin lands tails. This simplification is defensible on the view that we are considering, since even though it is possible that I should lose my entire fortune between now and the tossing, so that my last dollar would be very valuable, and even though it is possible that my competitor should welsh on his bet, it is under ordinary circumstances practically certain that neither of these possibilities will be realized.

Since probabilities are intervals, even if we assume that utilities are real valued, expected utilities must also be construed as intervals: The expected utility of P's being true is the interval comprised by the utility of P's being true, multiplied by the lower probability of P, and the utility of P's being true multiplied by the upper probability of P. If, relative to my corpus of practical certainties, the probability of P is [q,r], and my utility for P is V, the expected utility of P is [Vq,Vr].

Of course one cannot simply "maximize" an interval, so the principle of maximizing expected utility cannot be phrased in the usual way. But

one thing does seem clear, and that is that it is irrational to choose an action whose maximum expected utility is less than the minimum expected utility of some other action. This limits our choices, but in general need not pick out a unique one as "rational It might be that some further constraint could be imposed (for example: choose the action whose minimum expected utility is a maximum) but this seems to come into competition with contrary constraints (for example, choose the action whose maximum expected utility is a maximum). It thus seems sensible, at this point, to limit ourselves to the following principle of rational action:

Principle III: In a situation in which X's corpus of practical certainties is S_p, X ought rationally to reject any choice C_i there is a C_j whose minimum expected utility exceeeds the maximum expected utility of C_i.

For example, suppose that choice 1 yields A with probability [.1,.2] and not-A with probability [.8,.9] and that the utility of A is 10, and of not-A is -1. The expected utility of C_1 is [1,2] + [-.8,-.9] = [.2,1.1]. Similarly, suppose that C_2 has an expected utility of [1.5,2.3] and that C_3 has an expected utility of [0.8,5.0]. The rule then requires rejection of C_1, but does not legislate between C_2 and C_3.

6. Deductive Performance.

There are a number of studies that purport to show that people are deficient in their deductive performance or competence or both. Of course, few people are brilliant logicians, and even brilliant logicians cannot be faulted for not living up to all of their deductive commitments. Nobody is aware of all the theorems of set theory, though many people are committed to them. The studies of deductive performance therefore ordinarily concern the failure of subjects to make relatively simple deductive inferences. One quite robust deficiency in this regard, apparent in both simple and complex tasks, has been called "confirmation bias" -- the tendency of people to look for or take account of evidence confirming a hypothesis rather than evidence falsifying a hypothesis.

One group of experiments in which subjects were to formulate hypotheses, test them, modify them, and so on, under circumstances designed to simulate those in which real scientists work was reported by Mynatt et al., (1977) and (1978). In both series of experiments, the investigators discovered a "confirmational bias." The subjects were tempted (1977, p.89) to formulate a hypothesis that was incorrect to account for the initial data with which they were presented. They were subsequently given a choice between pairs of experimental setups, some of which could falsify that hypothesis, others of which could provide disconfirming evidence for it, or suggest alternative hypotheses, and others of which could only provide confirming data. Many of the subjects chose to examine evidence that could only provide confirming evidence for their initial hypothesis, even when they had received instructions to the effect that it was the job of the scientist to "disconfirm theories and hypotheses." "Subjects who

started with triangle hypotheses, regardless of [whether they were told to confirm or to disconfirm hypotheses] chose at a much higher than chance rate screens [presenting evidence] which could only confirm such hypotheses." (p. 93) On the other hand, "subjects could use falsifying data when they got it." (p. 94)

This suggests that the subjects were not (for the task at hand) deficient in deductive prowess, but deficient in strategic awareness when it came to testing universal generalizations. The second, more complex, series of experiments confirmed the results of the first, "even though the instructional manipulations were much more extensive than in the earlier study." (1978, p. 404) On the other hand, in this more complex task subjects did not almost always abandon disconfirmed hypotheses. The authors do not suggest deductive incompetence as the reason, but write: "the more complex environment may have made generation of new hypotheses, and hence, abandonment of disconfirming hypotheses, more difficult." In the second series, the subjects were explicitly told that the phenomena obeyed a uniform set of deterministic laws, but presumably that information was also implicit in the instructions to the first set of subjects.

It is quite clear that the authors began (and perhaps ended) with the conviction that some form of Popperian (Popper, 1959) approach to hypothesis testing was appropriate (Mynatt et al. (1977), p. 85), Mynatt et al. (1978), p. 396). By the end of the second series, apparently some doubts had been raised (Mynatt et al. (1978), p. 405). But even then, they have not abandoned the phrase "confirmation bias," even though they suggest that the "confirmation bias may not be completely counterproductive" (p. 405).

On the view sketched earlier, the confirmation "bias" is generally

perfectly appropriate. In the "natural ecology" (Einhorn and Hogarth (1981)) few universal generalizations present themselves as live possibilities; indeed, on the view suggested, universal generalizations can be maintained only by being made features of the scientific language. The basic inductions which can serve as a guide in life are statistical. It is obviously utterly irrelevant to look at non-B's for evidence concerning "%(A,B) ε [p,q]". This is so however close p may be to 1. One does not look among non-B's for evidence either supporting or controverting the statement "Almost all A's are B's." Only A's are relevant.

At a more sophisticated level of science -- note that this did not become a significant part of empirical knowledge until the last few hundred years -- one does encounter genuine universal generalizations. But even then, as Kuhn (1962) and Feyerabend (1970) have pointed out, such generalizations are, particularly early on, maintained in the face of falsifying evidence. That is, they are maintained "come what may"; the characteristic, as Quine (1961) has suggested of linguistic conventions.

Every student knows, of course, that there are universal generalizations in real science. Most students are told (alas) that they are empirical and falsifiable. But every student has also experienced the apparent falsification of an accepted universal law -- if only Archimedes' law in a physics laboratory -- only to be told that the reason for the apparent falsification is that he has "not done the experiment correctly" or "not interpreted the results correctly."

The students who were the subjects of the experiments of Mynatt et al. were therefore in a bind: they are torn between the natural rational attempt to confirm a hypothesis of the form "Almost all A's are B's," and the knowledge that there is a universal relation to be found, while at the

same time they know that universal relations in general survive apparent disconfirming instances. It is no wonder that they reported being frustrated (Mynatt et al.(1978), p. 404). Mynatt et al. also report ((1978), p. 405) that "the three subjects who quickly abandoned disconfirmed hypotheses did not rapidly progress."

If clever students are given a list of alternative hypotheses, one of which is known to be true, one would conjecture that with relative efficiency they would proceed to falsify hypotheses until there was but one left. If they are given the information that there is a useful pattern for prediction to be discovered, one would conjecture that they would (correctly, rationally) seek to confirm a statistical hypothesis of the form "Almost all A's are B's" or "Almost none of the A's are B's" by examining the A's. If they are told that there exists a universal hypothesis, but need to make up their own list, one would conjecture that they would oscillate between the two approaches: looking for useful clues to support a hypothesis of the form "Almost all A's are B's," and supporting it by looking for more A's; considering a finite and nonexhaustive list of hypotheses of the form "All A's are B's" and perhaps rejecting items in the list in the face of falsifying evidence, or perhaps modifying them. One would expect just such a confused picture as that revealed by the protocols of the experiments cited.

Closely related to the data just reviewed are the data provided by Wason (1960) and Johnson-Laird and Wason (1977). In these simpler experiments the subjects are presented with four cards showing A, D, 4, and 7. They are presented with the statement: If a card has a vowel on one side,

it has an even number on the other side. Their task is to determine which cards must be turned over in order to know whether the statement is true or false. The subjects do not do well; many say that the cards marked A and 4 should be turned over; some just mention card A. Few give the correct answer, which is A and 7.

Again, we seem to have encountered "confirmation bias." There is no doubt in this simple case that subjects are answering incorrectly. But there are several factors that can be called on to account for their mistakes. First, to account for the answer "A only," the previous suggestion may apply: In the "natural ecology" the generalizations with which people mainly deal, and the ones which they habitually confirm, are essentially statistical: "Almost all X's are Y's." For testing these generalizations, only observations of X's are appropriate. The natural tendency to treat a generalization, "If something is an X then it is a Y," as representing "Almost all X's are Y's," and to test it by looking only at X's, carries over to the artificial task in which the natural tendency is incorrect.

Second, it is not uncommon in ordinary English to use the conditional, "If something is X then it is Y" to express what is more accurately expressed by a biconditional. This comes about, I conjecture, because under many circumstances it is already known -- already an item in the corpus of practical certainties of both speaker and listener -- that Y's are X's. "If the hay is too wet, it will mold," will ordinarily (and correctly) be understood as having the same meaning as: The hay will mold if and only if it is too wet. Combined with the first tendency, this would lead subjects to examine both A and 4.

It is interesting to contrast the results of this experiment with the results of a "formally" similar experiment (Johnson-Laird *et al.*, 1972) cited by Johnson-Laird and Wason (1977). "The subjects were instructed to imagine that they were postal workers engaged in sorting letters on a conveying belt; their task was to determine whether the following rule had been violated: 'If a letter is sealed, then it has a 5*d* stamp on it.'" The material consisted of the back of a sealed envelope, the back of an unsealed envelope, the front of an envelope with a 5*d* stamp, and the front of an envelope with a 4*d* stamp. Almost all the subjects correctly chose to examine both the sealed envelope with the unseen face, and the envelope with the 4*d* stamp.

Cohen (1981, p. 324) suggests that the difference is due to "familiarity and concreteness in the letter sorting task." There may be elements of this, but a more salient distinction, in the framework being suggested here, is that the rule in the Postal example, is a genuine, stipulative, *a priori*, *rule*: All sealed letters *shall* have, *must* have 5*d* stamps. It is not the rule that is being tested, but the conformity of the letters to the rule. It would be interesting to test the performance of subjects on a similarly concrete and familiar task where the "rule" is not a stipulative one, but a descriptive generalization such as: If a letter has a 5*d* stamp, it has the return address on the back.

Some of the earliest work on the relation between logic and thinking was done by Mary Henle (1962). She presents evidence "that even where the thinking process results in error, it can often be shown that it does not violate the rules of the syllogism. Many errors were found to be accounted for not in terms of a breakdown of the deductive process

itself, but rather in terms of changes in the material from which the reasoning proceeds."(p. 377) The material used in her studies was deliberately chosen to be informal, and her subjects were (as far as possible) "logically naive." Under these circumstances it would be difficult indeed to ensure that the reasoning processes of the subjects did not use material from their own bodies of knowledge.

In ordinary argument, this dependence on a body of practical certainties is even more pronounced. Johnson-Laird (1977) offers the example (from Abelson and Reich, 1969): "He went to three drugstores, therefore the first two drugstores didn't have what he wanted." In such cases it is clear that the argument is not deductive: nobody's corpus of practical certainties excludes the possibility of the premise being true and the conclusion false. The argument is of the statistical "almost always" form. It is also clear that the persuasiveness of the argument, its probabilistic soundness, depends on two things that can be represented in the framework suggested. First, enough knowledge in the rational corpus (of both arguer and listener) to warrant the inclusion of the statistical statement: "Almost always when a person goes to three drugstores, it is because the first two didn't have what he wanted." (This is noted by Johnson-Laird.) And second (not remarked on by Johnson-Laird), knowledge of the drugstore visiting person which allows him to be a random member of the set of people who visit three drugstores with respect to having a particular reason for doing so. Consider the difference, for example, if the totally ambiguous "he" in the argument is replaced by "Tom" -- the argument still goes through -- and if it is replaced by the definite description, "the oldest comparison shopper employed by Rite-Aid" -- the argument fails.

7. Probabilistic Inference.

Some of the most recent and popular work on human inference making concerns the alleged deficiencies in the ability of people to perform probabilistic inference correctly. In itself, this is not surprising; statistical argument is more complex than deductive argument by its very nature. Indeed, its principles have yet to be formulated in a way that conforms to the intuitions of professional statisticians. (This raises a difficulty for the suggestion of Stitch and Nisbet (1980) that one should turn to "experts" for criteria of justification. In statistical inference there are large groups of acknowledged experts upholding contrary criteria of justification.)

One well known example (Kahneman and Tversky (1973)) concerns the "neglect of base rates." In this experiment subjects are told that in a certain town 85% of the cabs are blue, and 15% of the cabs are green. A witness to an accident identifies a cab as green, and it is given that under the circumstances he can make correct identifications of color 80% of the time. The subjects were then asked for the probability that the cab involved in the accident was blue.

The median estimated probability was 0.2. The authors claim that this shows a serious error, since the contingency table employing the general relative frequency of blue and green cabs looks like this:

	seen as blue	seen as green	
truly blue	.68	.17	.85
truly green	.03	.12	.15

The conditional probability that a cab is blue, given that the witness says it is green, is thus .17/(.17 + .12) = .59.

Cohen (1981) claims that this is no error at all -- that the subjects are right and the investigators wrong. "The fact that cab colours actually vary according to an 85/15 ratio is strictly irrelevant to this estimate because it neither raises nor lowers the probability of a specific cab-colour identification being correct on the condition that it is an identification by the witness. A probability that holds uniformly for each of a class of events because it is based on causal properties, such as the physiology of vision, cannot be altered by facts, such as chance distributions that have no causal efficacy in the individual events" (p. 328-329).

Cohen's argument seems wrong. Suppose the story were changed; suppose that it is given as a pure problem in inference. Suppose the cab in question were not singled out by having been in an accident, but was selected by some stochastically random procedure from among the cabs in the town. Otherwise the story is the same. Regardless of the "causal basis" of the witness's identification of the color of the cab, it is clear that the contingency table would give the correct probability that the cab is blue: .59.

An experiment like this was described by Lyon and Slovic (1976): The numbers are kept the same, but the population is a population of light-bulbs, of which 15% are defective, and the "witness" is a scanning device which is 80% accurate. The lightbulb in question was explicitly said to be chosen at random. Again, it was discovered in this experiment, as well as in a large variety of similar experiments, that the subjects tended to ignore the base rate, or to give it insufficient weight.

Nevertheless, there is a difference between the experimental results for the two problems, as reported by Lyon and Slovic. In a version of the cab

problem in which the probability of green was asked for, the median estimate was .80. In the corresponding lightbulb problem, the median estimate was also .80. But the interquartile range was different in the two problems: in the cab problem it was reported as .80-.80. In the lightbulb problem it was .25 - .80, where the correct answer is .41. This difference is revealing: it suggests that a significant number of the subjects in the lightbulb problem did attempt to take account of the base rate, while practically none of the subjects in the taxicab problem did so.

One possible explanation of this would lie in the fact that no information about the relative frequency with which blue and green cabs are involved in accidents is given in the problem. It would be interesting to see if the results were more like those of the lightbulb problem if it were stated that 15% of the cabs involved in accidents were green and 85% were blue. It would also be interesting to ask the subjects in the original taxicab problem to estimate the proportion of accidents involving cabs that involve blue cabs. Would it be the canonical 85%? Or would subjects say "there isn't enough information"? Or would subjects (improperly) infer from the one case they "know" about that blue cabs are less likely to be involved in accidents than green cabs?

It is worth observing that there is a reconstruction of the cab problem which strongly supports the intuitions of the subjects who ignore base rates. Suppose the corpus of practical certainties of the subject contains the following statistical knowledge

(1) %(Cabs,blue) = .85

(2) %(Cabs identified as green,blue) = .20

(3) %(Cabs in accidents identified as green,blue) ε [0,1]

(4) %(This particular cab in an accident identified as green,blue) ε [0,1]

These statistical statements mention increasingly specific potential reference sets. The relevant proportions in (1) and (2) differ, so the holder of this corpus should base his probability on (2) rather than on (1); (3) and (4) concern more specific reference sets yet, but they provide no new information. According to the rules for choosing reference classes (Kyburg, 1974), (2) provides the appropriate basis for the probability that the particular cab in question is blue.

It is also worth noting that this kind of reconstruction is not permissible in the lightbulb example. Corresponding to (3) we would have:

(3') %(selected lightbulbs testing defective,non-defective) [0,1]

Given the conditions of the problem, that the lightbulbs are selected at random, it follows from statements corresponding to (1) and (2) that

(3") %(selected lightbulbs testing defective,non-defective) = .59

The argument leading to (3") is arithmetically nontrivial for most people. It is thus not surprising that the subjects didn't get the right answer. But I would conjecture that it would be quite easy, with pencil and paper, to convince the subjects that .59 <u>was</u> the right answer. On the other hand, it is not so easy to convince all subjects that .59 is the right answer in the cab problem; L. J. Cohen, for example, who has access to unlimited supplies of paper and pencils, is still unconvinced (1979, 1980, 1981).

There is still the fact to be explained that the median answer in the lightbulb problem is 0.80; many subjects must have answered the lightbulb problem exactly on the lines of the cab problem. A hypothetical explanation of this might run as follows: In assessing probabilities in ordinary life, one can ordinarily use a frequency in a reference set. People have relatively

little experience in combining probabilities in accordance with Bayes' Theorem. For example, it seems more natural in giving the probability of a one on a toss of a die, given that it is an odd number, to calculate that a third of the odd numbered tosses yield a one, than to divide a sixth (ones) by a half (odd numbers). Perhaps this is something that could be got at by a cleverly designed experiment.

Given a choice of conflicting frequencies in two possible reference sets, therefore, it is usually the case that one of those frequencies will be suitable and the other unsuitable as a basis for a probability. In the lightbulb problem the correct reference set is not one of the options given: the subject must devise the correct reference set, and compute its relevant/relative frequency, on his own. The difficulty that subjects have with the lightbulb problem, therefore, seems to stem from what might be called the natural ecology of multiple choice questions. Given the matrix of relative frequencies as part of the data in the lightbulb problem, how would subjects do?

There are several other biasses in intuitive probabilistic inference that have been subjected to experimental assessment. Three such biasses are discussed in a well known paper by Tversky and Kahneman (1974). The representativeness heuristic leads to bias in the assessment of the probability that an A is a B: If A and B are very much alike, the probability of a (particular) A being a B is assessed as higher than it should be; if A and B are very dissimilar, the assessed probability is excessively low. The representativeness heuristic leads to the neglect of prior probabilities. In one particular experiment described in Tversky and Kahneman (1974) and reported in Kahneman and

Tversky (1973), the subjects were told the appropriate base rates, but neglected them: " subjects evaluated the likelihood that a particular description belonged to an engineer rather than to a lawyer by the degree to which this description was representative of the two stereotypes with little or no regard for the prior probabilities of the categories" (Johnson-Laird and Wason (1977), p. 328).

It seems clear that in this case the subjects are flatly wrong. It is curious that when a description was given that could fit either of the categories equally well, the subjects still ignored the base rates, taking the probabilities to be .5 and .5. A possible explanation of this is that the subjects acted as if the base rate among the individuals selected to be categorized was 50%. It is not hard to imagine subjects beguiling themselves as follows: The individual selected is either a lawyer or an engineer; that's one of each, so the base rate among those selected is .50, and the only relevant evidence I have to decide between the two alternatives consists of the description.

Even more blatant violations of normative statistical theory are to be found when representativeness is applied to frequencies, either in estimating the frequency in a sample from a known population or in estimating the population from which a sample of known frequency has been drawn (Tversky and Kahneman, 1974). Here again, the subjects are simply in error; but a possible explanation is at hand. Statistical generalizations of the forms "almost all A's are B's, and "Almost no A's are B's" are the sorts of generalizations most people usually find most useful in their dealings with the real approximate world. But the representativeness heuristic does not work badly with respect to generalizations of this sort.

The authors offer other conjectures as to why people fail generally to learn from experience statistical facts concerning regression, or the relation between sample size and variability. No doubt many factors are at work. But it should not be concluded -- as one might conclude if the competence of ordinary people were taken as a standard of rational belief -- that because few people take account of regression to the mean in making predictions statistical theory does not provide a norm of rationality.

On the other hand, there are many instances of intuitive probabilistic inference in which it is not clear how to apply statistical theory. In another paper (Kahneman and Tversky 1979), the same authors offer some suggestions for improving prediction. They emphasize the importance of considering "distributional" information, as opposed to relying too heavily on the information embodied in the unique case under consideration. "The analyst should therefore make every effort to frame the forecasting problem so as to facilitate the utilization of all the distributional information that is available to the expert" (typescript, p. 5). Since the authors accept a subjectivistic interpretation of probability, they accept that individuals may assign probabilities to particular cases that are not based on any form of statistical knowledge. Their emphasis on distributional knowledge, from our point of view, reflects a logical truism: There can be <u>no</u> (rational) probabilities without underlying frequencies.

Of course, there are all kinds of distributional information. The unique case under consideration can be seen to fall in a great many classes about which we have statistical information. Thus it is crucial for practical guidance, as well as for the characterization of rationality,

to be able to sort out that distributional information. In the author's simple case of a publisher attempting to predict the sales of a book, they suggest first that "the selection of a reference class is straightforward," (p. 8) and, later, that "For example, the reference class for the prediction of the sales of a book could consist of other books by the same author, or books on the same topic, or of books of the same general type ... the most inclusive class may allow for the best estimate of the distribution of outcomes, but it may be too heterogeneous to permit a meaningful comparison to the book at hand ... the class of books on the same topic could be the most appropriate" (p. 9).

The authors give no normative criteria for the choice of a reference class, but it is clear that this is a problem that is on their minds. On the view of probability and rationality being developed here, it is clearly crucial. It is a matter that receives considerable formal attention in the full development of epistemological probability, but it would take us too far afield to consider it in detail here. As Einhorn and Hogarth (1981, p. 65) point out, "There is no generally accepted normative way of defining the appropriate population."

8. Decision Under Uncertainty.

In their review of behavioral decision theory, Einhorn and Hogarth (1981) draw attention to a number of apparent conflicts between the ordinary normative theory (subjective expected utility theory) and the behavior of real agents. They point out that even nonregressive estimates may turn out to be more profitable than regressive ones in an environment that is nonstationary. "[T]he optimal prediction is conditional on which hypothesis you [the experimenter] hold." All of the oddities of intuitive

probabilistic inference are reflected in the differences between the recommendations of normative choice theory and the description of the ways in which people choose. But there are other differences as well. For example, the choice problem may be stated in two apparently equivalent ways, and/lead to different choices under each statement. Figure/ground relations, learning, attention, etc., all play a role, in addition to the role played by utility and probability, in the explanation of human choice. As the authors say (p. 75), " the descriptive adequacy of $\underline{E}(\underline{U})$ [expected utility theory] has been challenged repeatedly.

Furthermore, the normative adequacy of $\underline{E}(\underline{U})$ has itself begun to be challenged, for example, by prospect theory. A particularly telling challenge on an intuitive level is provided by Lopes (1981). She discusses the traditional St. Petersburg paradox, and a piece of anecdotal evidence reported by Samuelson (1963).

The St. Petersburg paradox goes like this: A fair coin is tossed until heads first appears -- say on the nth toss. The player then receives a prize of $2^{\underline{n}}$ dollars -- or, to avoid questions of the utility of money, $2^{\underline{n}}$ utiles. What is the fair price for the player to pay for the privilege of playing the game once? The answer -- the expected value of the game -- turns out to be infinite. Most people would pay relatively little. Is this irrational?

Lopes turns the problem around: If someone offers to run the game for a number of players, only charging $\underline{k}$ dollars, is he necessarily irrational? Using dollar units, a number of Monte Carlo simulations[6] were run in which 100 businesses, starting with a capital of $10,000, sell opportunities to play the Petersberg game at prices of $25, $50,

and $100. After a million customers, the prospects of the businesses selling chances for $100 are not bad: only 10% of them had gone broke, and the mean and median outcomes were 56 and 79 million dollars respectively. Not bad for a business with only a $10,000 start-up cost. The only problem is finding enough customers. It is not correct to say (as Lopes does, p. 378) that if it is a good business for the businessman, it cannot be a good one for the customers -- most successful businesses survive because both the businessman and the customer increase their utilities in the exchange -- but it does seem unlikely that the Petersberg business can compete successfully with the State Lottery.

Within the framework at hand, the analysis of the game is quite straightforward. Suppose the index of practical certainty is .001. In the evidential corpus it is assumed known that the coin is fair; it is thus practically certain that the game will last no more than ten tosses, and its practical expected value will be $10.00. The entrepreneurs will have a hard time selling chances for $100.00.

How does this differ from the State Lottery? One should be practically certain that one is not going to win the lottery; is one therefore irrational in buying a ticket? As is well known, there are circumstances in which it is not irrational. They depend on the relation between utility and money, and the fact that money is not of monotonically increasing utility. A dollar isn't worth much -- what can you do with a dollar? -- but having a fortune would be very nice. In buying a lottery ticket, your practical expected value should be 0 -- you should be practically certain that you won't win. On the other hand, the marginal disutility of parting with a single dollar can also be 0, so in terms of practical expected utility the

exchange is fair for you. If you count into the practical expected utility the opportunity to daydream about winning, the exchange is better than fair.

Another situation discussed by Lopes comes from Samuelson (1963). Samuelson tells of offering to bet some colleagues $100 to $200 that a specified side of a coin would not appear on its first toss. None of the colleagues took him up. One person argued as follows (quoted from Lopes, p. 382):

> I won't bet because I would feel the $100 loss more than the $200 gain. But I'll take you on if you promise to let me make 100 such bets ... One toss is not enough to make it reasonably sure that the law of averages will turn out in my favor. But in a hundred tosses of a coin, the law of large numbers will make it a darn good bet. I am, so to speak, virtually sure to come out ahead in such a sequence...

Samuelson finds this response irrational. Lopes sides with the colleague. The colleague feels, correctly, that he can be practically certain that in a series of a hundred gambles, he will come out ahead. It is true that he may not; he may lose $10,000. But the probability of this is very small -- lower than the probability that on his next air trip he will be killed. A practical man does not take such possibilities as real. (Then why buy air insurance? The explanation is roughly the same as that for buying a lottery ticket. You'll hardly miss a couple of dollars, and although your practical expectation is 0 or slightly negative in dollars, you obtain the added utility of peace of mind.)

Lester Dubins and L. J. Savage (1965) provide a subjectivistic account of a structurally similar situation. Suppose you have $1000 and absolutely

have to have $10,000 by the next day. You are in a casino, and gambling is your only hope. How should you do it? The answer is clearly that you should stake the whole $1000 on a single 10:1 shot. The more you divide your stake, the more probable it is that the house odds will get you.

An attempt to make sense not only of the choices that people actually make when faced with uncertainty but of relatively clear and compelling intuitions concerning such choices, is prospect theory (Tversky and Kahneman 1981). It is not clear to me exactly how well it accords with the views presented here, but there are clearly certain similarities. In this theory decision weights correspond to "subjective probabilities," but they are different in a number of respects. They do not sum to 1 (subcertainty), just as the relevant probabilities in the St. Petersberg problem do not sum to 1. "The function is not well behaved near the endpoints" (p. 454). On the present account, probabilities greater than p, the level of practical certainty, are treated as 1, and probabilities less than 1-p are treated as 0. The asymmetry between large and small probabilities in prospect theory is not reflected here. But it may be that this apparent asymmetry is more fruitfully taken account of in the valuation function, about which I have nothing to say here.

9. Conclusion:

These reflections suggest a three-fold connection between the philosophical normative investigation of rationality, and the empirical psychological study of belief.

The first connection is that empirical studies may suggest certain facts relevant to the development of normative constraints. The normative constraints must be appropriate to the kinds of beings we are. This is

not to say that we must automatically embody them, or even that we must be able to achieve them (a counsel of perfection), but that we must be capable of approaching them, of learning to do better. At the same time, most empirical studies take for granted certain normative constraints. (In studying degrees of belief, we assume that people's bodies of knowledge are consistent, for example.) These presupposed constraints may or may not be appropriate; if they are inappropriate, they may vitiate the results of the psychological investigation.

The second connection is that philosophical investigations into rationality may provide a useful framework within which psychological investigation can be conducted. For example, a structure which allows for some kind of probabilistic acceptance may prove useful for exploring the oddities of choice under uncertainty when that uncertainty is reflected by chances close to 0 or close to 1.

The third and most obvious connection is provided by the fact that whatever we wish to conclude about the rationality with which people draw conclusions or apportion their beliefs, we want our own conclusions to be rational, to be well supported by the evidence. Even an argument to the effect that (other) people's beliefs do not conform to sound inductive canons ought itself to be based on sound inductive principles.

To sum up: I take a theory of rational belief to be a normative theory. I take its object to be the improvement of our understanding. While such a theory cannot be made up out of whole cloth -- it must be appropriate to the beings whose understanding we are trying to improve -- neither should it merely reflect what people actually do. Intuition and introspection and empirical investigation may reveal general principles in simple and concrete

cases. Analysis and argument may reveal connections among these principles, or defects in them as applied to more complicated cases, or limitations in their scope.

What I have tried to do here is to illustrate this process by showing the way in which my approach to probability and inductive acceptance throws light on several things:

(a) The deductive structure of the set of rationally accepted statements (assuming there is one), thus providing a connection between deductive cogency and rational belief that is lacking (or only implicit) in logic itself.

(b) The addition and deletion of statements to and from a body of accepted statements; this is a matter that involves probability, but it is one that pure Bayesian conditionalization can throw no light on.

(c) Degrees of belief in statements that are not accepted. I suggest that the constraints are more extensive than Bayesians often suppose (requiring conformity to statistical knowledge), but less precise (being represented by intervals, rather than real numbers).

(d) A distinction between conditional degrees of belief (reflected in betting ratios on conditional bets) and degrees of belief conditional on the acceptance of new data.

These matters constitute a single tangled web -- there is no way in which we can approach them piecemeal. And since the web is so tangled, we have an explanation both of the inconclusiveness of psychological experiments designed to explore rationality of belief, and of the frustration that has led some philosophers to despair of finding a "broad reflective equilibrium" of rationality (Cohen 1981). But though I think it is clear

that the problem of characterizing rationality is a difficult one -- far more difficult than many have realized -- it does not seem insuperable. The very difficulties we uncover contribute to our understanding. And the fact that we progress at all, the fact that we listen to each other's arguments, and recognize an obligation to deal with them, suggests that our goal can be approached.

FOOTNOTES

1. Some remarks on notation may be helpful. I use capital letters P, Q, R, etc., to stand for declarative sentences in the object language: "Fido is in the manger," "There is a black dog in the manger," "All the dogs on the farm are in the manger," "All crows are black," "Between 50% and 70% of the successful conceptions yield brown offspring." The capital letter S I reserve for the set of sentences that constitutes the background knowledge, or the body of reasonably accepted beliefs, or the rational corpus, of the agent. Lower case letters x, y, z, etc., are metalinguistic variables, representing terms of the object language: These may be names of individuals ("Fido"), sets ("the set of crows"), sets of sets ("the set of subsets of the set of crows"), etc. Relative logical type, which is all we need be concerned with, is given by context. For example, "$x = y$" and "$x \subset y$" tell you that x and y are of the same logical type; "$x \in y$" tells you that whatever type x may be, y is of the type of sets that have as members objects of type x. This flexibility is essential if we want to handle both the probability that the next counter is black, and the probability that the set of counters we have examined is representative of the proportion of black counters in the bag. The lower case letters, p, q, r, etc., are used both as metalinguistic variables taking as values real number designators in some canonical form (binary, decimal) and also the real numbers between 0 and 1 so designated. Thus [p,q] may represent the expression "[.6,.7]" in the object language, and may also represent, in our metalanguage, the closed interval of real numbers [.6,.7].

2. It follows from the principle of epistemic conditionalization that if S is relevant to T, then T is relevant to S, where S is relevant to T just in case the degree of belief in T, given S, $B_S(T)$, is different from the unconditional belief in T, B(T). Suppose that S is the statement that in the long run about 60% of the tosses of this coin land heads, and T is the statement that the next toss of this coin lands heads. Take the degree of belief in T, relative to our ordinary knowledge of coins, to be 0.5. S is relevant to T:

$$B_S(T) = 0.6 = \frac{B(S \wedge T)}{B(S)} \neq B(T).$$

But it is surely stretching this to say that T is relevant to S. That a coin has been tossed and landed heads shouldn't change the probability of the statistical statement:

$$B_T(S) = B(S)$$

3. Or too strong, if there are no non-probabilistic dispositions to act.

4. I believe so many things to be true that I am almost certain that at least one of them must be false.

5. This has been argued at length in Kyburg (1961), Kyburg (1974), and a number of papers, many of which are collected in Kyburg (1983a).

6. Monte Carlo simulations involve programming a computer to undertake a vast number of trials embodying computer-selected random numbers to determine the outcomes of the trials. The computer simulation thus can reflect the long run outcome of a stochastic process.

References

Abelson, R.P. & Reich, C.M. (1969) Implication modules: a method for extracting meaning from input sentences. First International Joint Conference on Artificial Intelligence, Washington, D.C., p. 41.

Bar-Hillel, Y. (ed.) (1965) Logic, methodology, and philosophy of science. Amsterdam: North-Holland.

Carnap, R. (1950) The logical foundations of probability. Chicago: University of Chicago Press.

Cohen, L.J. (1977) The probable and the provable. Oxford: Oxford University Press.

(1979) On the psychology of prediction: whose fallacy? Cognition 7:385-407.

(1980) Whose is the fallacy? Cognition 8:89-92.

(1981) Can human rationality be experimentally demonstrated? The Behavioral and Brain Sciences 4:317-331.

De Finetti, B. (1980) Foresight: Its logical laws, its subjective sources. In Studies in subjective probability, ed. H. E. Kyburg & H. Smokler, p. 53-118. Huntington, N.Y.: Krieger.

Dempster, A. (1968) A generalization of Bayesian inference. Journal of the Royal Statistical Society Series B, 205-247.

Dretske, F. (1981) Knowledge and the flow of information. Cambridge, Mass.: MIT Press.

(1983) Precis of Dretske's knowledge and the flow of information. The Behavioral and Brain Sciences

Dubins, L. & Savage, L.J. (1965) How to gamble if you must. New York: McGraw Hill.

Edwards, W. (1975) Comment. Journal of the American Statistical Association 70:291-293.

Edwards, W. (1954) The theory of decision making. Psychological Bulletin 51:380-417.

Einhorn, H.J. (1978) Confidence in judgment: the persistance of the illusion of validity. Psychological Review 85:395-416.

(1980) Learning from experience and suboptimal rules in decision making. Cognitive Processes in Choice and Decision Behavior, ed. T. S. Wallsten, 1-20. Hillsdale N.J.: Erlbaum.

Einhorn, H.J. & Hogarth, R.M. (1981) Behavioral decision theory: processes of judgment and choice. Annual Reviews of Psychology 32:53-88.

Ellis, B. (1979) Rational belief systems. Oxford: Blackwell.

Falmagne, R.J. (ed.) (1975) Reasoning, representation, and process. Hillsdale, NJ: Erlbaum.

Feyerabend, P. (1970) Against method. In Studies in the philosophy of science, ed. Radner & Winokur, 17-130. Minneapolis: Univ. of Minnesota Press.

Finetti, Bruno De (1973) Bayesianism: its unifying role for both the foundations and the application. Proceedings of the 39th Session of the International Statistical Institute.

Good, I.J. (1962) Subjective probability as a measure of a non-measureable set. In Logic, methodology and philosophy of science, eds. Nagel, Suppes & Tarski, 319-329. Stanford: Stanford University Press.

Hempel, C.G. & Oppenheim, P. (1945) A definition of 'degree of confirmation'. Philosophy of science 12:98-115.

Henle, M. (1962) On the relation between logic and thinking. Psychological Review 69:366-378.

Hintikka, J. (1965) Towards a theory of inductive generalization. In Logic, methodology, and philosophy of science, ed. Y. Bar-Hillel. Amsterdam: North-Holland.

Jeffrey, R.C. (1965) The logic of decision. New York: McGraw Hill.

(1974) Preference among preferences. Journal of Philosophy 71:377-391.

Johnson-Laird, P.N. & Wason, P.C. (1977) A theoretical analysis of insight into a reasoning task. In Thinking, eds. Johnson-Laird & Wason, 143-157.

Johnson-Laird, P.N. (1977) Reasoning with quantifiers. In Thinking, eds. Johnson-Laird & Wason, 129-142.

Johnson-Laird, P.N. and Wason, P.C. (1977) Thinking: Readings in Cognitive Science. Cambridge: Cambridge University Press.

Kahneman, D. & Tversky, A. (1972a) On the psychology of prediction. Oregon Research Institute Bulletin 12, University of Oregon.

(1972b) Subjective probability: a judgment of representativeness. Cognitive Psychology 3:430-454.

(1973) On the psychology of prediction. Psychological Review 80:237-251.

(1979a) On the interpretation of intuitive probability: a reply to Jonathan Cohen. Cognition 7:409-411.

(1979b) Intuitive prediction: biases and corrective procedures. Management Sciences 12:313-327.

Keynes, J.M.(1952) A treatise on probability. London: Macmillan & Co.

Korner, S. (ed.)(1957) The Colston papers. London: Butterworth's Scientific Publications.

Kuhn, T.S. (1962) The structure of scientific revolutions. Chicago: University of Chicago Press.

Kyburg, H.E. Jr. (1961) Probability and the logic of rational belief. Middletown, CT.: Wesleyan University Press.

(1974) The logical foundations of statistical inference. Dordrecht: Reidel.

(1983a) Epistemology and inference. Minneapolis: University of Minnesota Press.

(1983b) Theory and measurement. Cambridge: Cambridge University Press.

Levi, I. (1974) On indeterminate probabilities. Journal of Philosophy 71:391-418.

(1980) The enterprise of knowledge. Cambridge, MA: The MIT Press.

Lopes, L.L. (1979) Doing the impossible: a note on induction and the experience of randomness, Department of Psychology, University of Wisconsin, Madison.

(1981) Decision making in the short run. Journal of Experimental Psychology 7:377-385.

Lyon, D. & Slovic, P. (1976) Dominance of accuracy information and neglect of base rates in probability estimation. Acta Psychologica 40:287-298.

Mellor, D.H. (1971) The matter of chance. Cambridge: Cambridge University Press.

(1980) Prospects for pragmatism. Cambridge: Cambridge University Press.

Mises, R. von (1957) Probability statistics and truth. London: George Allen and Unwin.

Mynatt, C.R., Doherty, M.E. & Tweeney, R.D. (1977) Confirmation bias in a simulated research environment: an experimental study of scientific inference. Quarterly Journal of Experimental Psychology 29:85-95.

(1978) Consequences of confirmation and disconfirmation in a simulated research environment. Quarterly Journal of Experimental Psychology 30:395-406.

Nagel, E., Suppes, P., & Tarski, A. (eds.) (1962) Logic, methodology and philosophy of science. Stanford, CA: Stanford University Press.

Nisbett, R. & Ross, L. (1980) Human inference: strategies and shortcomings of social judgement. Englewood Cliffs: Prentice-Hall.

Popper, K.R. (1957) The propensity interpretation of the calculus of probability and the quantum theory. In The Colston Papers, ed. Korner, 65-70.

(1959) The logic of scientific discovery. London: Hutchinson & Co.

Quine, W.V.O. (1960) Word and object. New York: John Wiley & Sons.

Radner, M. & Winokur, S. (eds.) (1970) Studies in the Philosophy of Science. Minneapolis: University of Minnesota Press.

Reichenbach, H. (1949) The theory of probability. Berkeley and Los Angeles: University of California Press.

Russell, B. (1948) Human knowledge. New York: Simon & Schuster.

Samuelson, P.A. (1963) Risk and uncertainty: a fallacy of large numbers. Scientia 98:108-113.

Savage, L.J. (1954) The foundations of statistics. New York: John Wiley & Sons.

(1966) Implications of personal probability for induction. Journal of Philosophy 58:593-607.

Shafer, G. (1976) A mathematical theory of evidence. Princeton: Princeton University Press.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1977) Behavioral decision theory. Annual Reviews of Psychology 28:1-39.

Smith, C.A.B. (1961) Consistency in statistical inference and decision. Journal of the Royal Statistical Society Series B 23:1-37.

Stich, S. & Nisbett, R. (1980) Justification and the psychology of human reasoning. Philosophy of Science 47:188-202.

Tversky, A. & Kahneman, D. (1974) The belief in the 'law of small numbers'. Psychological Bulletin 76:105-110.

(1973) Availability: a heuristic for judging frequency and probability. Cognitive Psychology 5:207-232.

(1974) Judgement under uncertainty: heuristics and biases. Science 185:1124-1131.

(1981) The framing of decisions and the psychology of choice. Science 211:453-458.

Wallsten, T.S. (ed.) (1980) Cognitive processes in choice and decision behavior. Hillsdale, NJ: Erlbaum.

Wason, P.C. (1960) On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology 20:273-281.

(1977a) Self-contradictions. In Thinking, eds. Johnson-Laird & Wason, 114-128. Cambridge: Cambridge University Press.

(1977b) On the failure to eliminate hypotheses... a second look. In Thinking, eds. Johnson-Laird and Wason, 307-314. Cambridge: Cambridge University Press.

Whitehead, A.N. & Russell, B. (1950) Principa Mathematica. Cambridge: Cambridge University Press. (First edition, 1910).

Cognitive Science Technical Report List

University of Rochester

1. Steven L. Small and Margery M. Lucas. Word expert parsing: A computer model of sentence comprehension. May, 1983.

2. Gary S. Dell. The representation of serial order in speech: Evidence from repeated phoneme effect in speech errors. May, 1983.

3. Henry E. Kyburg, Jr. Rational Belief. May, 1983.

4. Alan H. Schoenfeld. Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. May 1983.

5. David A. Taylor, Jung-Oh Kim, and Padmanabhan Sudevan. Representation of linear orders. May, 1983

6. Alan H. Schoenfeld. Episodes and executive decisions in mathematical problem solving. May, 1983.

7. Alan H. Schoenfeld. The wild, wild, wild, wild world of problem solving (a review of sorts). May, 1983

8. Ellen Carni and Lucia A. French. The acquisition of before and after reconsidered: What develops? June, 1983.

9. Susan M. Garnsey and Gary S. Dell. Some neurolinguistic implications of prearticulatory editing in production. July, 1983.

10. Patrick J. Hayes. The second naive physics manifesto. October, 1983.

Note: Requests for U of R Cognitive Science Reports should be addressed to either the author or to Cognitive Science Reports, Department of Philosophy, University of Rochester, Rochester, NY 14627.